PHIL 7001: Homework Assignment 1

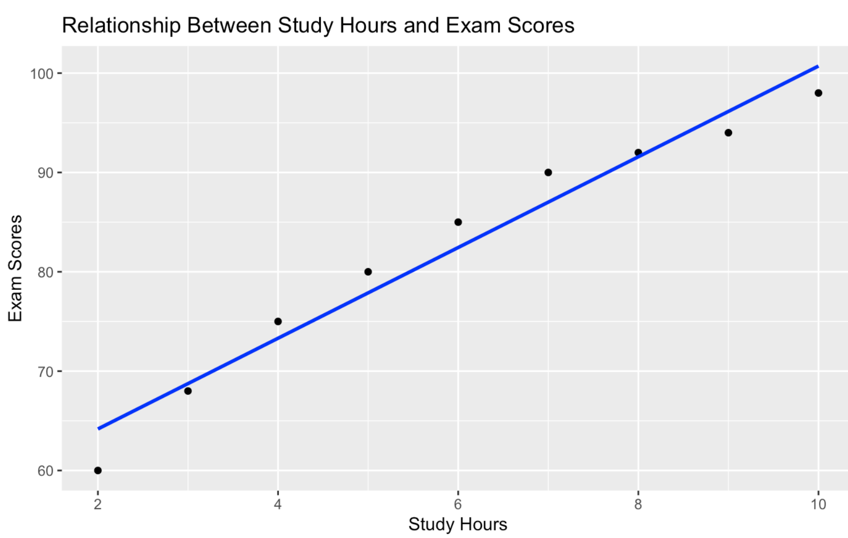
Professor Boris Babic

Problem 1: Introduction to R (10 points)

```
# Load the ggplot2 library if not already loaded
library(ggplot2)

# Create the data frame
data_students <- data.frame(
  study_hours = c(2, 4, 3, 5, 7, 6, 8, 9, 10),
  exam_scores = c(60, 75, 68, 80, 90, 85, 92, 94, 98)
)

# Create the scatter plot with a regression line
ggplot(data_students, aes(x = study_hours, y = exam_scores)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "blue") +
  labs(x = "Study Hours", y = "Exam Scores",
       title = "Relationship Between Study Hours and Exam Scores")
```



Problem 2: Probability Basics (20 points)

a. Conditions for a probability function: $0 \leq P(A) \leq 1$, $P(\Omega) = 1$, and $P(A \cup B) = P(A) + P(B)$ for disjoint events.

b. The vector of probability assignments (0.6, 0.8, 0.6) is illegitimate because the sum of the probabilities assignments for events A, B, and C exceeds 1. To make it a legitimate assignment of probabilities, we need to scale down the probabilities proportionally. This can be done by dividing each probability by the sum of all probabilities in the vector, which will be (0.3, 0.4, 0.3).

c. $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.3 + 0.5 - 0.1 = 0.7$

d. Probability of neither A nor B: $P((A \cup B)') = 1 - P(A \cup B) = 1 - (0.6 + 0.8 - 0.5) = 0.1$

e. $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.15}{0.3} = 0.5$

Problem 3: Bayes' Rule (20 points)

Step-by-step answers:

Let's define the events:

- Event D: The patient has the disease (the flu).
- Event T: The patient received a positive test result for the flu diagnosis.

We want to find the conditional probability $P(D|T)$, which can be calculated using Bayes' Rule:

$$P(D|T) = \frac{P(T|D) \cdot P(D)}{P(T)}$$

Given information:

- Prior probability of the flu occurrence $P(D)$: 0.01 (1%)
- Probability of getting a positive test results given that the patient has the disease $P(T|D)$: 0.96 (sensitivity)
- False Positive rate $P(T|\neg D)$: 0.03 (1 - specificity)

To calculate the marginal likelihood $P(T)$, we need to consider two cases: the patient has the disease (D) and the patient not having the disease ($\neg D$):

$$P(T) = P(T|D) \cdot P(D) + P(T|\neg D) \cdot P(\neg D)$$

Where:

$$P(\neg D) = 1 - P(D) = 0.99$$

Now, we can calculate $P(T)$:

$$P(T) = 0.96 \cdot 0.01 + 0.03 \cdot 0.99 = 0.0393$$

Finally, plug these values into Bayes' Rule:

$$P(D|T) = \frac{0.96 \cdot 0.01}{0.0393} \approx 0.2443$$

So, the probability that a bulb is actually defective, given that it fails the test, is approximately 0.2443 or 24.43%.

Problem 4: Probability Spaces (25 Points)

a. The binomial distribution in mathematical form is:

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$$

Where X and k are random variables representing the number of successes; n is a parameter representing the number of trials; p is a parameter representing the probability of success.

b.

$$P(X = 3) = \binom{6}{3} \cdot (0.7^3) \cdot (0.3^3) = 0.1852 \text{ or } 18.52\%$$

c. The normal distribution in mathematical form is:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Where X is a random variable representing the values of the variable being examined; μ is a parameter representing the mean; σ is a parameter representing the variance; π and e are constants.

d.

```
1 - pnorm(65, mean = 50, sd = 10)

[1]0.0668072

or
```

```
pnorm(65, mean = 50, sd = 10, lower.tail = FALSE)
```

```
[1]0.0668072
```

e. Based on the empirical rule,

$$P(x > 14) = \frac{1}{2} \cdot (1 - 0.9544) \approx 0.0228$$

Problem 5: Linear Regression (25 Points)

a.

```
model <- lm(exam_scores ~ study_hours, data = data_students)
```

b.

```
summary(model)
```

c. Yes, the study hour is a statistically significant predictor of a student's exam score. Based on the model summary, the p-value associated with the coefficient of the predictor 'study_hours' is 3.99e-06, which is much smaller than the 5% significance level.

d. 5.99e-08. Thus, we have strong evidence to reject the null hypothesis that $\beta_0 = 0$.

e. Predicted exam score = 55.0444 + 4.5667*3 = 68.7445 $\approx$ 69